

Experiences of open science & software

Peter Bankhead

Senior Lecturer in Digital Pathology

University of Edinburgh

11/2018

What is Bioimage analysis?

Biology



Not what I do

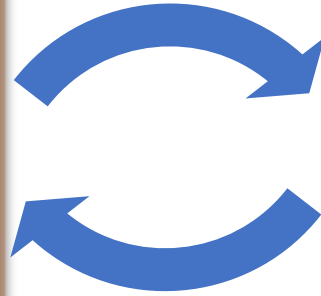


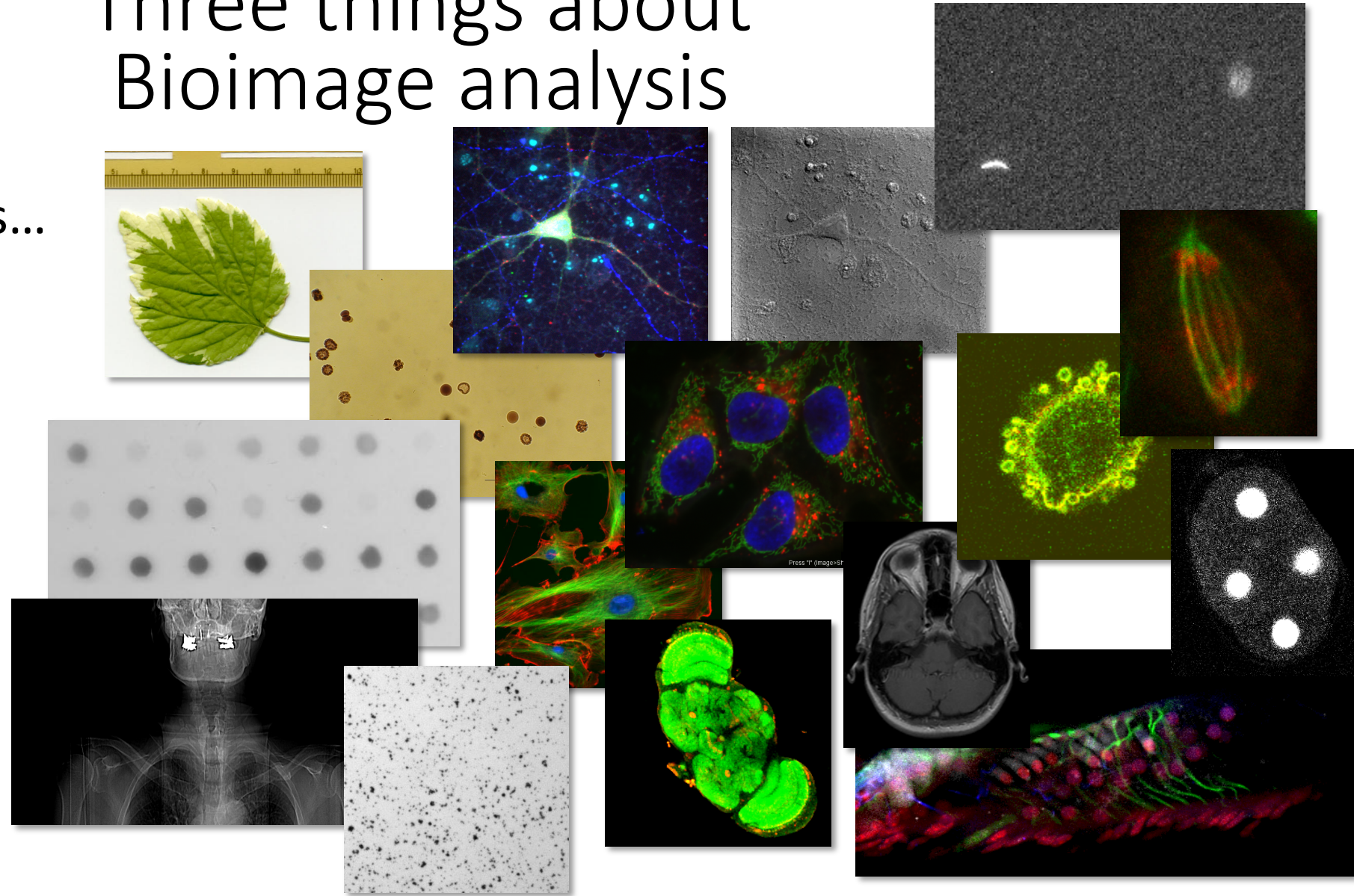
Image analysis



What I do

Three things about Bioimage analysis

- ...is diverse

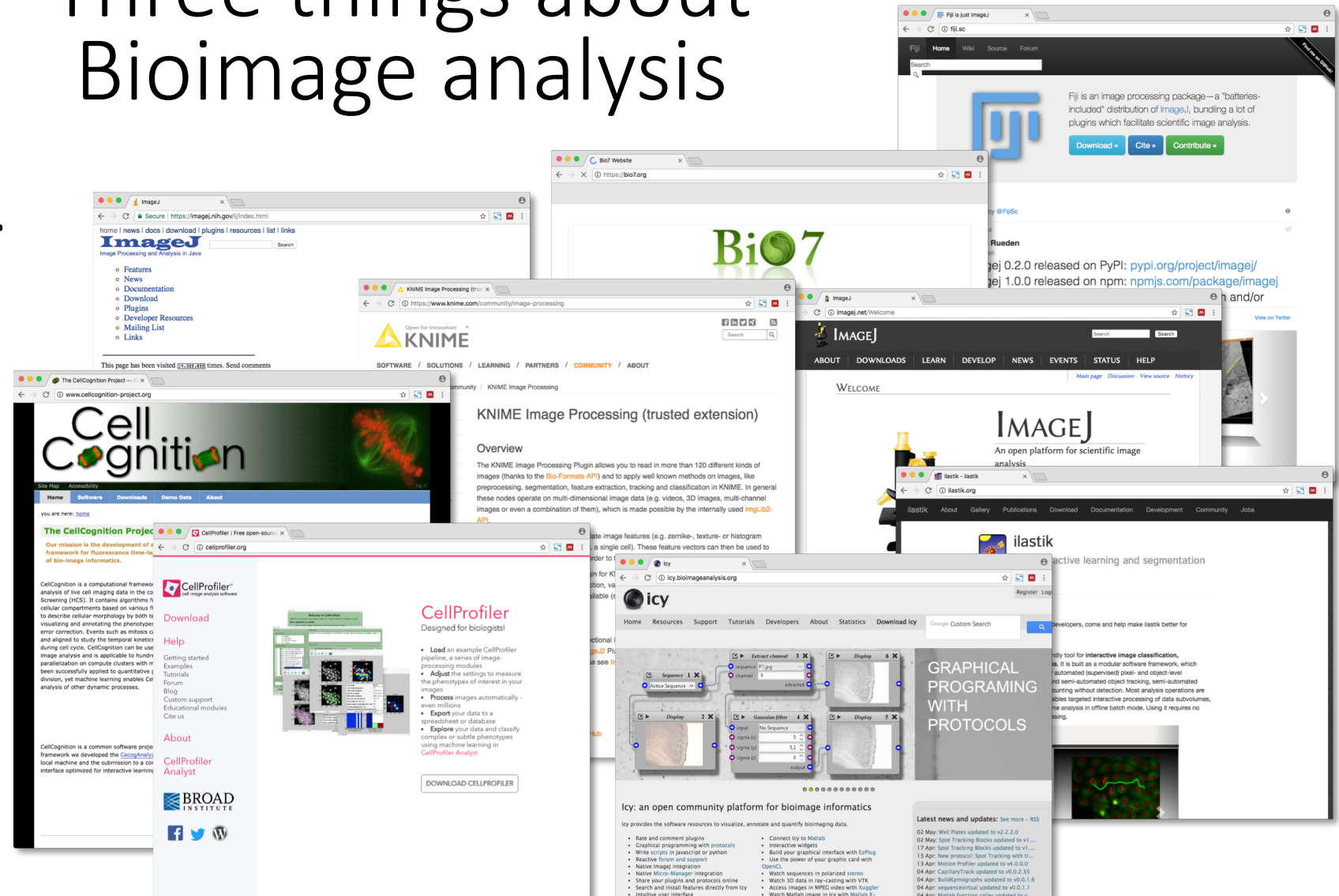


Sample images from *ImageJ* & *Fiji*

Three things about Bioimage analysis

Bioimage analysis...

- ...is diverse



Popular open source bioimage analysis software

Three things about Bioimage analysis

Bioimage analysis...

- ...is diverse
- ...is cross-disciplinary

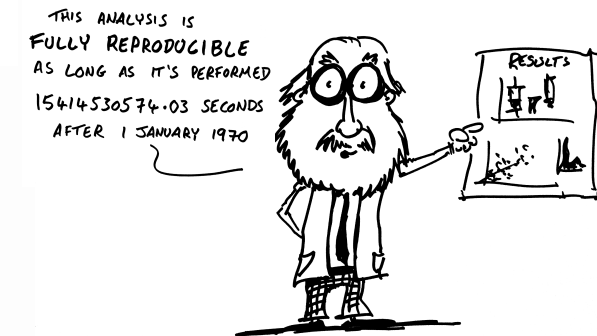
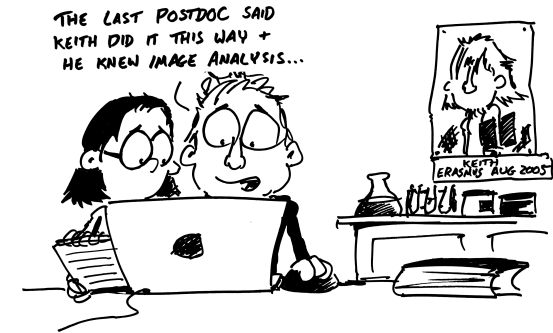
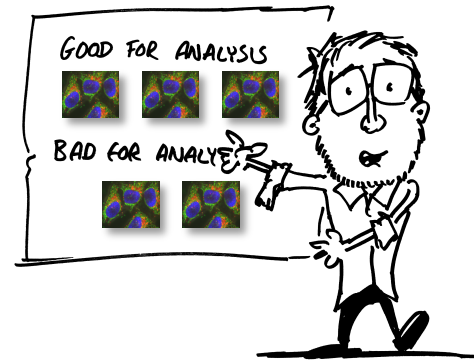


Sometimes the same person masters both skillsets –
but more often we need to work together

Three things about Bioimage analysis

Bioimage analysis...

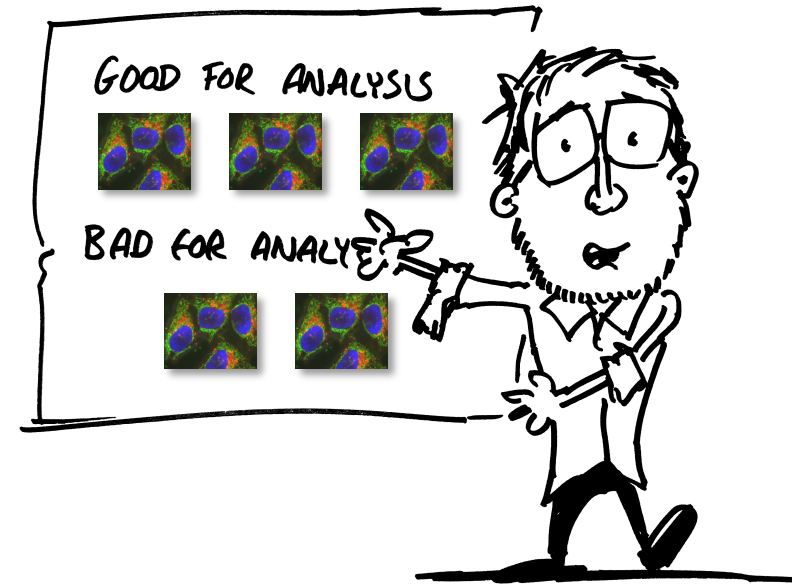
- ...is diverse
- ...is cross-disciplinary
- ...is often misunderstood



Three things about Bioimage analysis

Bioimage analysis...

- ...is diverse
- ...is cross-disciplinary
- ...is often misunderstood



Not all images are suitable for
quantitative analysis

Three things about Bioimage analysis

Bioimage analysis...

- ...is diverse
- ...is cross-disciplinary
- ...is often misunderstood



Just because someone used a method before doesn't mean it's right

Three things about Bioimage analysis

Bioimage analysis...

- ...is diverse
- ...is cross-disciplinary
- ...is often misunderstood

TO AVOID MANUAL BIAS,
WE TRIED ALL 17 AUTOMATED
THRESHOLDS IN IMAGES
AND CHOSE THE ONE THAT
WORKED



Image analysis is *not* unbiased
(even if a computer does it)

Three things about Bioimage analysis

Bioimage analysis...

- ...is diverse
- ...is cross-disciplinary
- ...is often misunderstood

THIS ANALYSIS IS
FULLY REPRODUCIBLE
AS LONG AS IT'S PERFORMED
15414530574.03 SECONDS
AFTER 1 JANUARY 1970

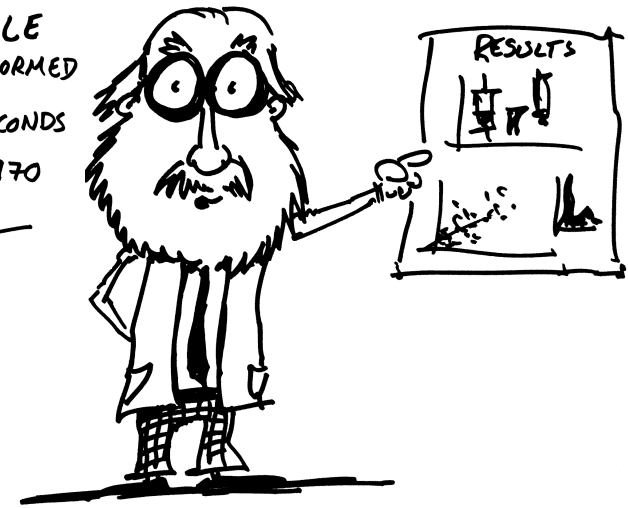
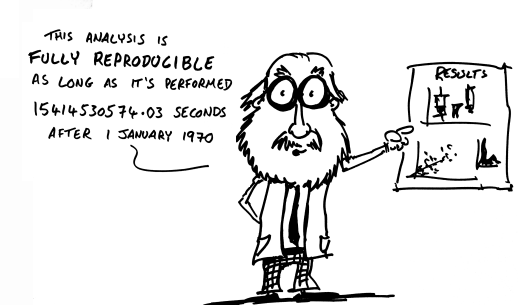
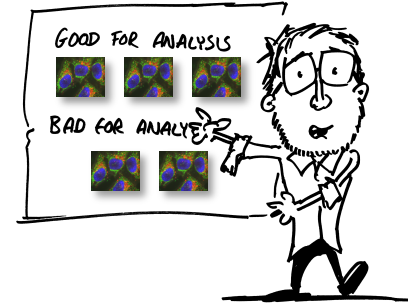


Image analysis *should* be reproducible –
but *it might not be!*

Open data & methods are essential!

We need to know *exactly*:

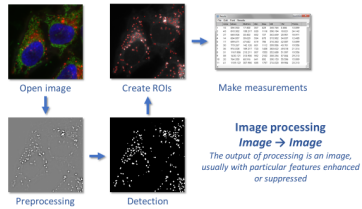
- What was analysed *Image repositories*
- How it was analysed *Open software*
- What were the results *Numbers & visualizations*



*Not every assumption will be recognized or reported –
we must be able to go to the data & analyse in a different way*

In Heidelberg, I chose to focus on teaching

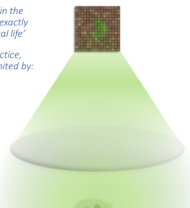
Image analysis converts an image into measurements & usually requires *image processing* to get there



Ideally structures in the image would map exactly to structures in 'real life'

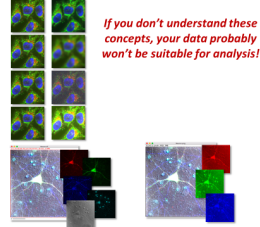
However, in practice, image quality is limited by:

- Blur
- Pixel size
- Noise



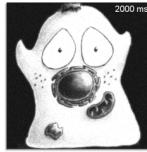
Five key concepts

1. Pixel size
2. Blur
3. Noise
4. Bit-depth
5. Colour

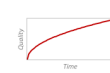


If you don't understand these concepts, your data probably won't be suitable for analysis!

Detecting more photons reduces noise



When the original number of photons is very low, detecting a few more dramatically reduces noise

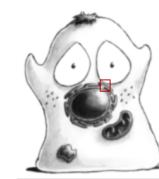


So could we get a better image just by increasing the exposure time?

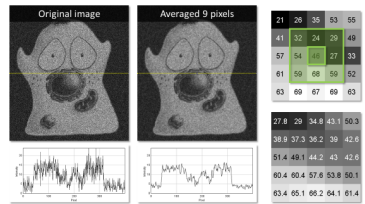
No!

Firstly, we could damage what we're looking at...
Secondly, we could saturate the image

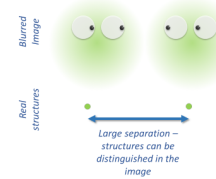
If these points are replaced by Airy disks, the result is a blurred image



Averaging adjacent pixels can reduce the noise in an image



Spatial resolution is a measure of how close two structures can be & yet still be distinguishable

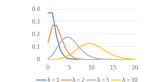


The number of emitted photons is random, but follows a Poisson distribution

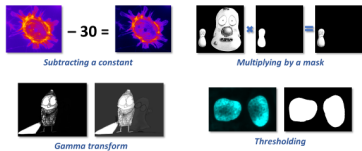


$$P(k; \lambda) = \frac{e^{-\lambda} \lambda^k}{k!}$$

λ = the true rate of photon emission
 k = an actual number of photons we might detect



We have already started image processing with *point operations*



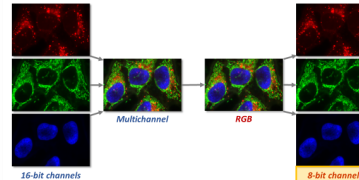
We will now look at more image processing with *neighbourhood operations*

Nonlinear point operations can also be used to enhance contrast – *but be very careful & always say if you use them!*

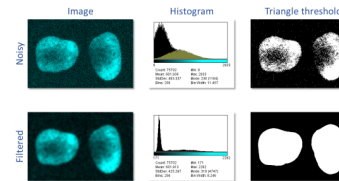


Gamma transform (Process → Math → Gamma...)
Pixel values p are replaced by p^r
(probably with extra scaling according to bit-depth)

When converting a multichannel image to RGB, information is usually lost



Smoothing an image can reduce noise – and sometimes greatly improves thresholding



Smoothing used Process → Filters → Gaussian Blur...

Exploring noise standard deviations

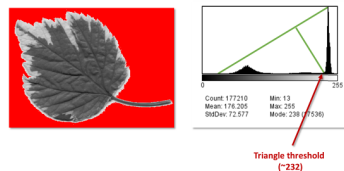
- Open the image *Happy_cell.tif*
- Test adding noise of different standard deviations with *Add Specified Noise...*

Question:

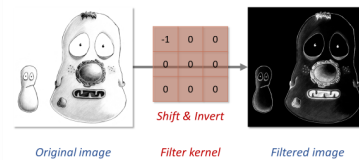
Does the brightness of the noise-free image make a difference to how bad the final appearance is after noise is added?

Test this by scaling the pixel values with the *Multiply* command before adding the noise

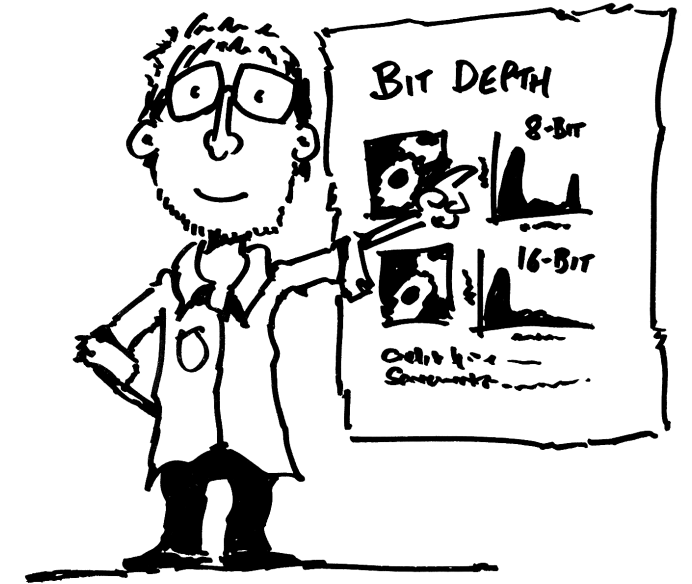
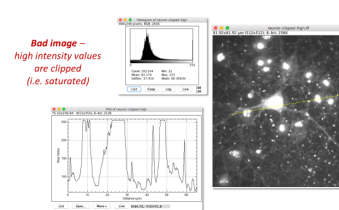
The triangle threshold assumes one large peak and uses the histogram shape



Other *linear* filters can be created by changing the weights of the *filter kernel*



Higher bit-depths help reduce the risk of *saturation & clipping*



This was useful, but its impact was limited by time & proximity...

Analyzing fluorescence microscopy images with ImageJ

Analyzing fluorescence microscopy images with ImageJ

Peter Bankhead
Queen's University Belfast

May 2014

This work is made available in the hope it will help a few more people develop an interest in image analysis. If you have any comments, corrections or suggestions, please contact me at p.bankhead@qub.ac.uk, so that it might one day get better.

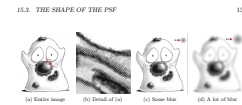


Figure 15.2: Images can be viewed as composed of small points (a), even if these points are not visible without high magnification (c). This allows us a useful way to understand what the response to a virtual image can be. Each point can be thought of as a small, different blur, the PSF. Images appear more or less blurred depending upon how large the blurry PSFs are (c) and (d).



Figure 15.3: Simplified diagram to help visualize how a light-emitting point would be imaged using a widefield microscope. Some of the light originating from the point is lost (the light being directed towards a focal point, the PSF is placed close to the focal point, the point-like detector would be small and single. However, if the light is placed close to the focal point, the intensity of the image is high. If you increase the size of the PSF, the intensity of the image is low.

15.3 THE SHAPE OF THE PSF

how great our blurring is. Blurring, it ends up being focused to some large volume known as the PSF, which has a minimum size dependent upon both the light's wavelength and the lens being used (Section 15.4.1).

The lensless, more particularly relevant if we consider that any fluorescent sample can be viewed as composed of many similar, exceedingly small light-emitting points – you may think of the fluorescent. Our image would ideally then include individual points too, digitized into pixels with values proportional to the emitted light. But what we get instead is an image in which every pixel has been replaced by its PSF, scaled according to the point's brightness. Where there are PSFs overlap, the detected light intensities are simply added together. Exactly how this looks depends upon the size of the PSF (Figure 15.2).

Section 15.3.3 gives an description of convolution as replacing each pixel in an image with a scaled filter – which is just the same process. Therefore it is no coincidence that blurring, because it comes extra short, characterizes the 1 about how blurring.

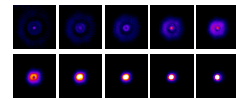


Figure 15.5: This shows from a central region of a fluorescent bead, starting from above and moving down to the focal plane. The same beam control settings have been applied to each slice as they are compared, although this causes the intensity band to appear somewhat more obvious; the image would not be visible at all. Because the image is (approximately) represented along the z-axis, additional slices moving below the focal plane would appear similar.

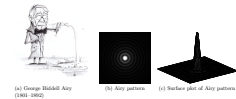


Figure 15.6: George Balaban and the Airy pattern. (a) During his school days, Airy had been interested for long while in the construction of point sources and other such devices (see <http://www.balaban.org/airy/airy.html>). The image surrounding the Airy disk has been blurred to the right in a panel. Although the image phenomenon was already known, Airy wrote the first theoretical treatment of it in 1835 (http://www.wikipedia.org/wiki/Airy_disk). (b) The Airy pattern, viewed as an image to which the constant has been set to reproduce the appearance of the central region surrounding the Airy disk. (c) A surface plot of an Airy pattern, which shows that the brightness is much higher within the central region when compared to the rings.

15.3 THE SHAPE OF THE PSF

how great our blurring is. Blurring, it ends up being focused to some large volume known as the PSF, which has a minimum size dependent upon both the light's wavelength and the lens being used (Section 15.4.1).

The lensless, more particularly relevant if we consider that any fluorescent sample can be viewed as composed of many similar, exceedingly small light-emitting points – you may think of the fluorescent. Our image would ideally then include individual points too, digitized into pixels with values proportional to the emitted light. But what we get instead is an image in which every pixel has been replaced by its PSF, scaled according to the point's brightness. Where there are PSFs overlap, the detected light intensities are simply added together. Exactly how this looks depends upon the size of the PSF (Figure 15.2).

Section 15.3.3 gives an description of convolution as replacing each pixel in an image with a scaled filter – which is just the same process. Therefore it is no coincidence that blurring, because it comes extra short, characterizes the 1 about how blurring.

15.3 THE SHAPE OF THE PSF

how great our blurring is. Blurring, it ends up being focused to some large volume known as the PSF, which has a minimum size dependent upon both the light's wavelength and the lens being used (Section 15.4.1).

The lensless, more particularly relevant if we consider that any fluorescent sample can be viewed as composed of many similar, exceedingly small light-emitting points – you may think of the fluorescent. Our image would ideally then include individual points too, digitized into pixels with values proportional to the emitted light. But what we get instead is an image in which every pixel has been replaced by its PSF, scaled according to the point's brightness. Where there are PSFs overlap, the detected light intensities are simply added together. Exactly how this looks depends upon the size of the PSF (Figure 15.2).

Section 15.3.3 gives an description of convolution as replacing each pixel in an image with a scaled filter – which is just the same process. Therefore it is no coincidence that blurring, because it comes extra short, characterizes the 1 about how blurring.

Section 15.3.3 gives an description of convolution as replacing each pixel in an image with a scaled filter – which is just the same process. Therefore it is no coincidence that blurring, because it comes extra short, characterizes the 1 about how blurring.

Section 15.3.3 gives an description of convolution as replacing each pixel in an image with a scaled filter – which is just the same process. Therefore it is no coincidence that blurring, because it comes extra short, characterizes the 1 about how blurring.

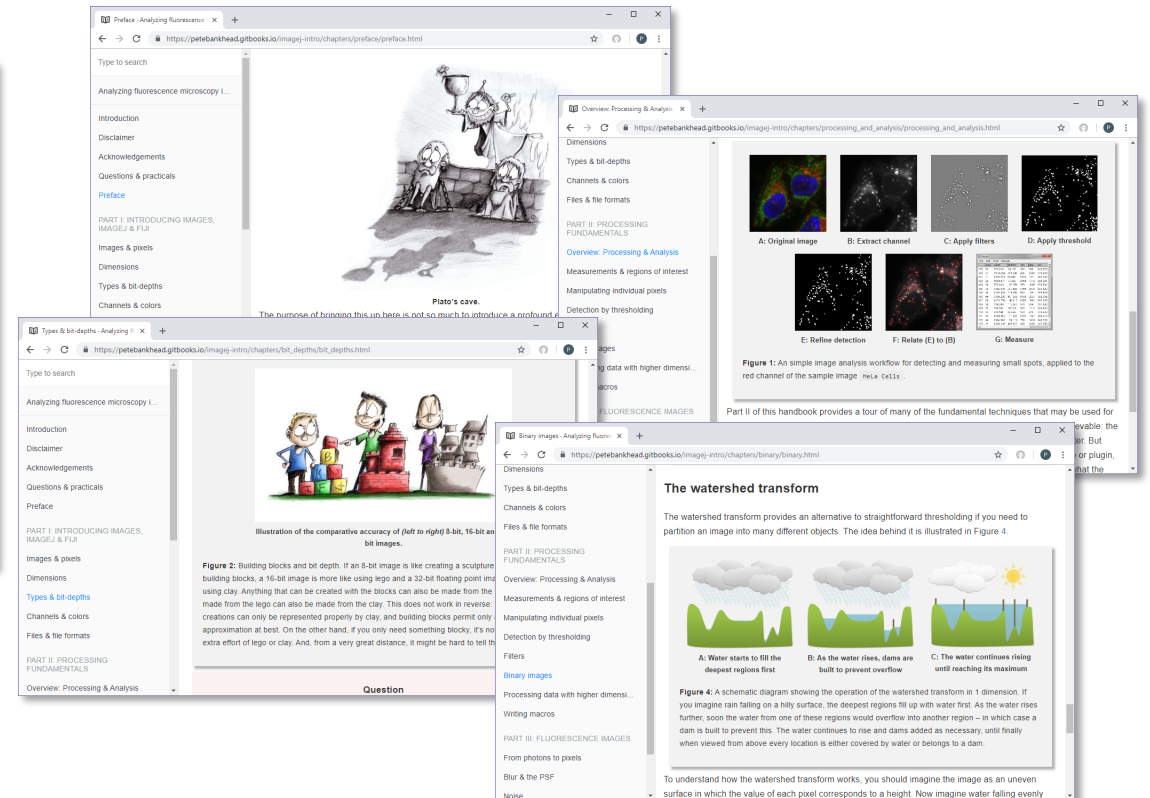
Section 15.3.3 gives an description of convolution as replacing each pixel in an image with a scaled filter – which is just the same process. Therefore it is no coincidence that blurring, because it comes extra short, characterizes the 1 about how blurring.

Section 15.3.3 gives an description of convolution as replacing each pixel in an image with a scaled filter – which is just the same process. Therefore it is no coincidence that blurring, because it comes extra short, characterizes the 1 about how blurring.

Section 15.3.3 gives an description of convolution as replacing each pixel in an image with a scaled filter – which is just the same process. Therefore it is no coincidence that blurring, because it comes extra short, characterizes the 1 about how blurring.

Section 15.3.3 gives an description of convolution as replacing each pixel in an image with a scaled filter – which is just the same process. Therefore it is no coincidence that blurring, because it comes extra short, characterizes the 1 about how blurring.

Section 15.3.3 gives an description of convolution as replacing each pixel in an image with a scaled filter – which is just the same process. Therefore it is no coincidence that blurring, because it comes extra short, characterizes the 1 about how blurring.



PDF version

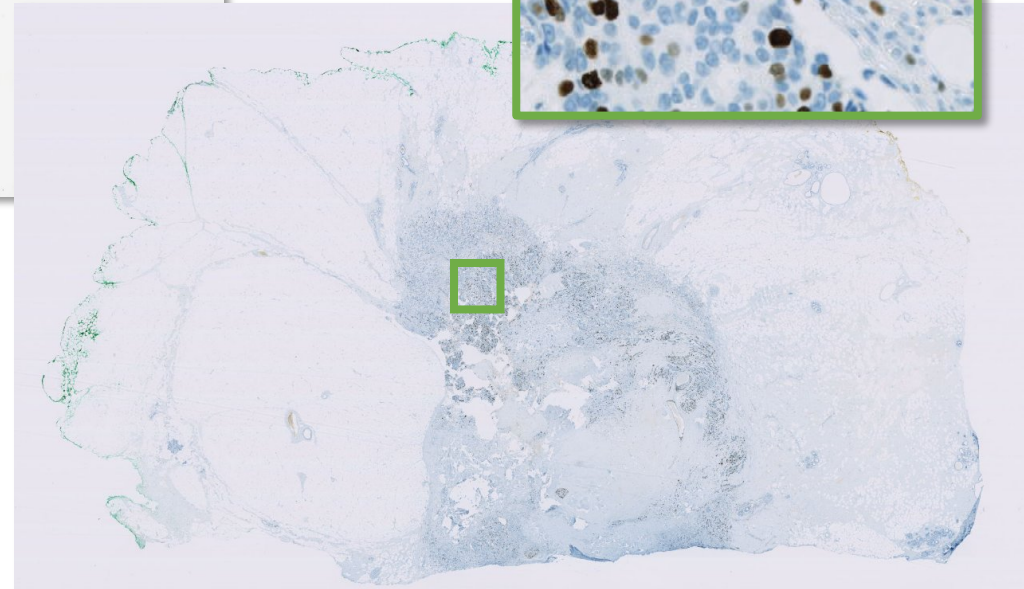
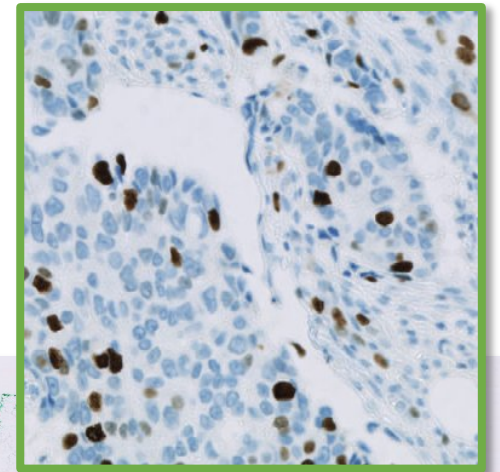
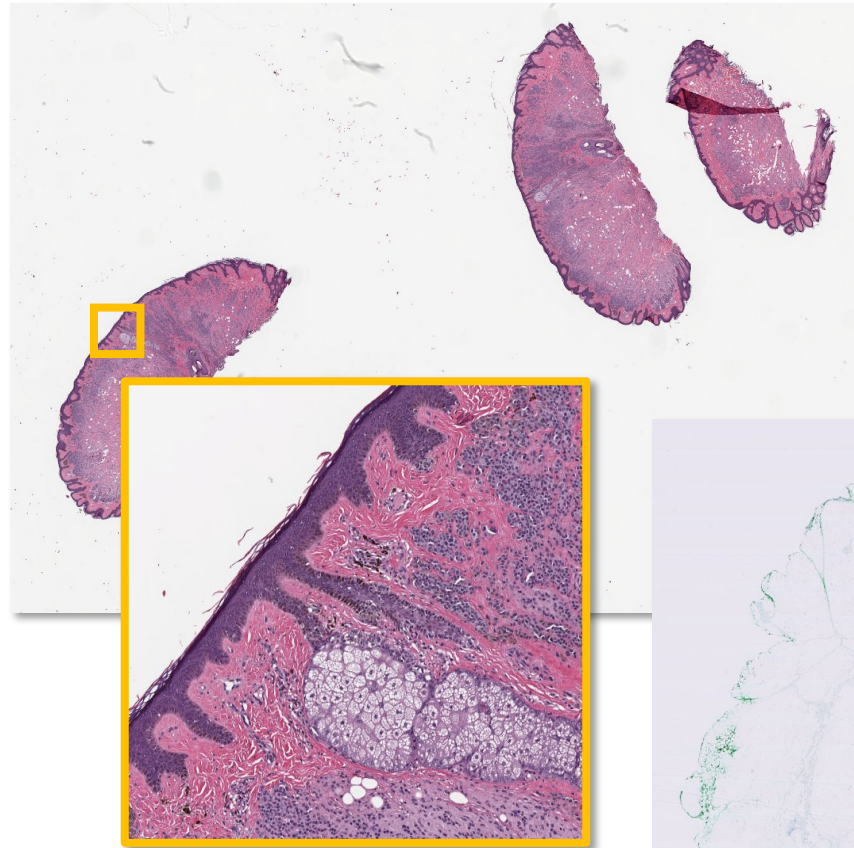
GitBook version

<https://petebankhead.gitbooks.io/imagej-intro>

New kinds of images require new kinds of software

Sometimes the necessary tools just don't exist

Pathology bridges science & medicine and plays a crucial role in patient care



Sample images from <https://openslide.org>

There is a critical shortage of pathologists

TESTING TIMES TO COME? AN EVALUATION OF PATHOLOGY CAPACITY ACROSS THE UK

NOVEMBER 2016

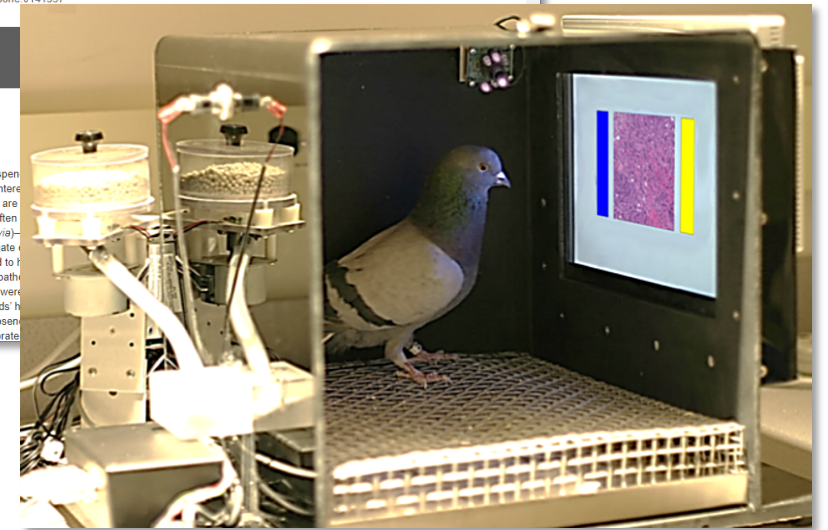
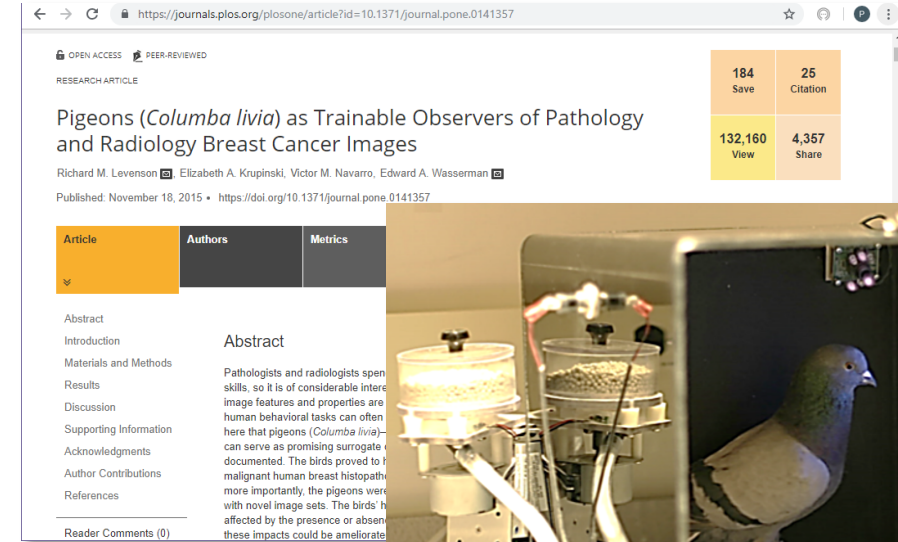
Pathology plays a major role in the diagnosis and treatment of cancer, as well as many other conditions. Pathology is comprised of 19 different disciplines and our research focussed on the most relevant to cancer: cellular pathology (which encompasses both histopathology and cytopathology); blood sciences; and molecular pathology.

Based on the number of pathologists currently in training and the age profile of the current workforce, our study found there is likely to be a severe crisis in pathology capacity within the next five to ten years.

There are two main approaches to address the needs in pathology



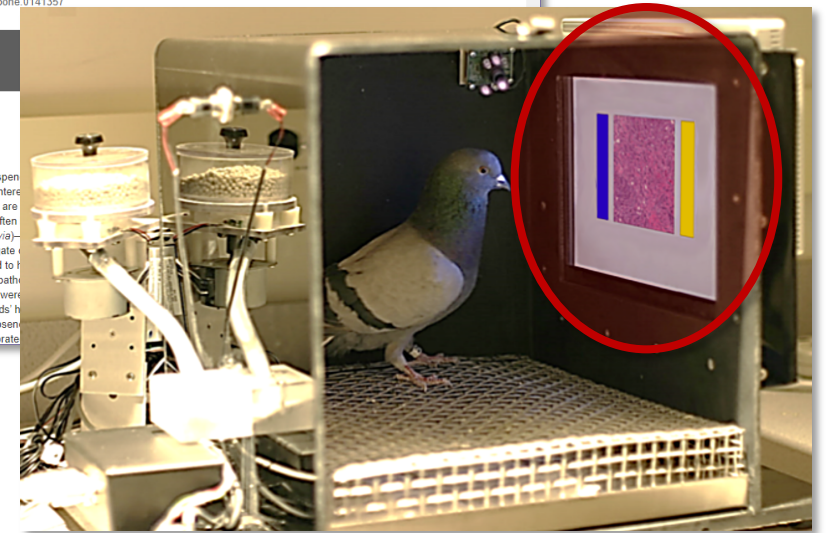
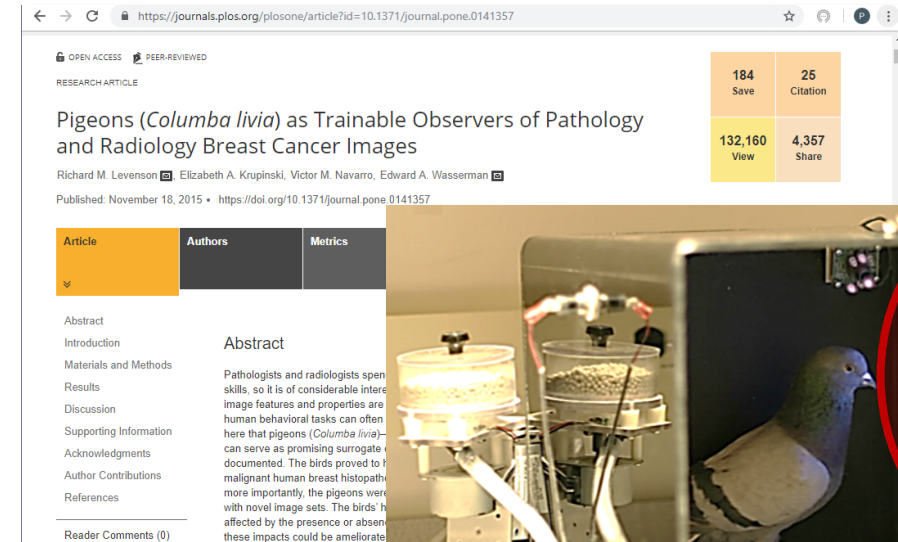
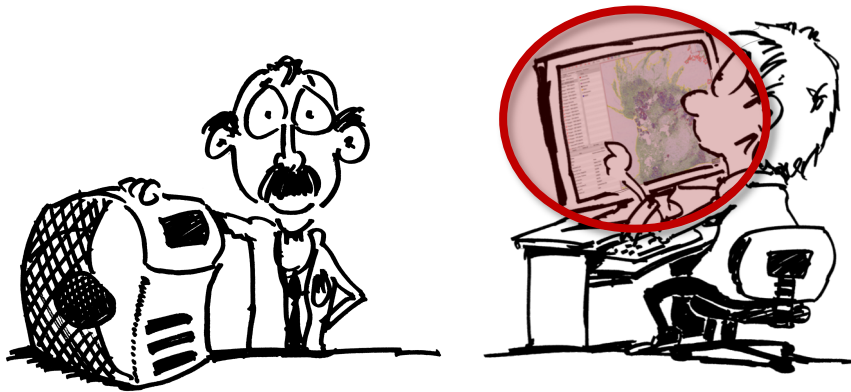
Digital pathology



Levenson RM, et al. PLoS ONE (2015)

Pigeons

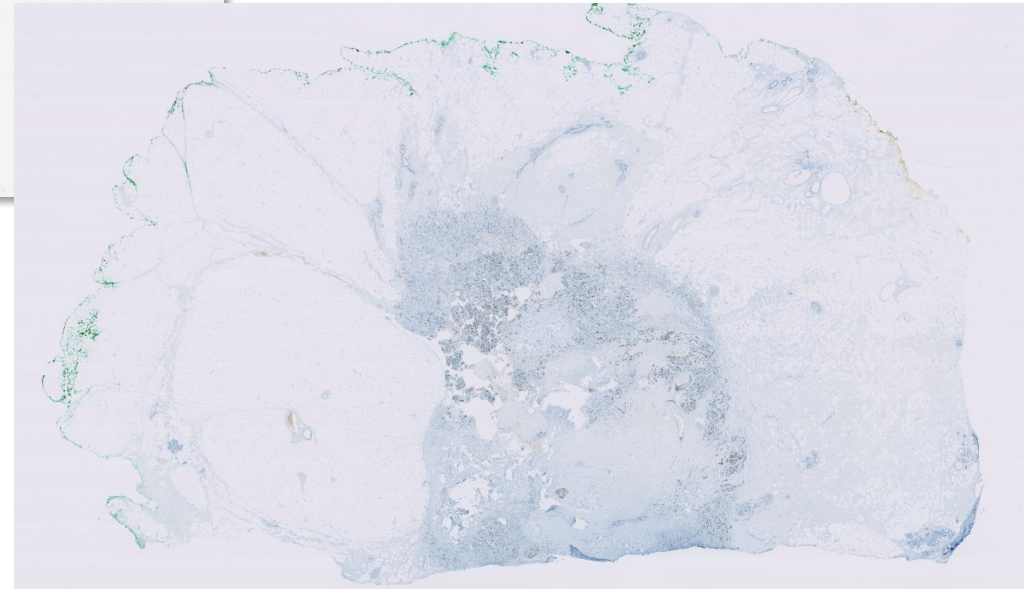
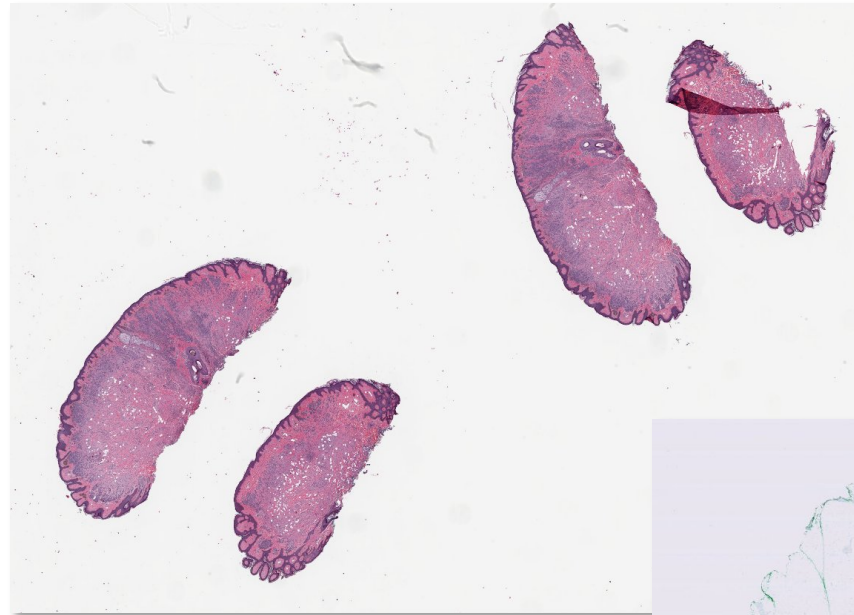
There is *really only one* main approach to address the needs in pathology



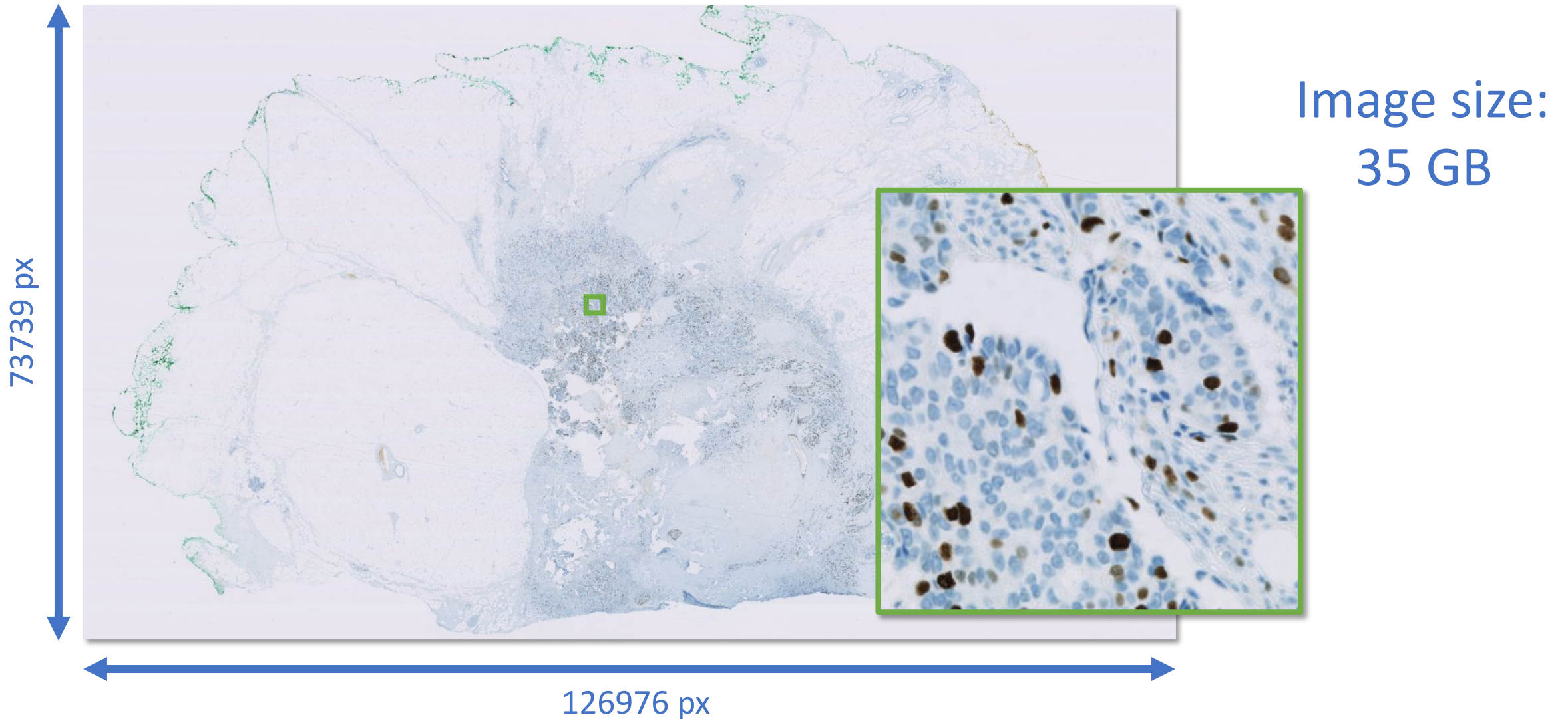
Levenson RM, *et al.* PLoS ONE (2015)

Digital pathology

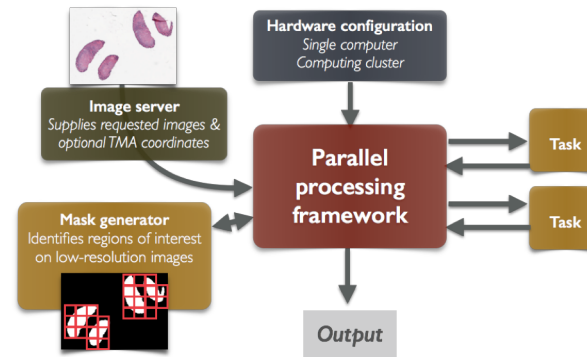
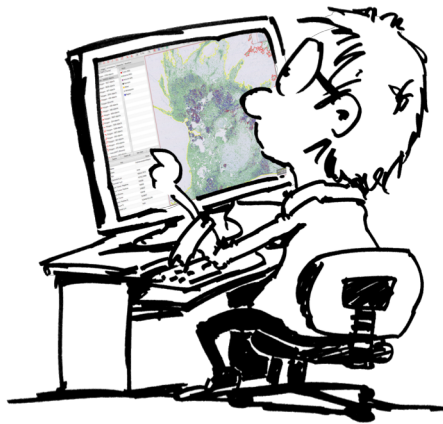
Digital pathology is based on
whole slide images



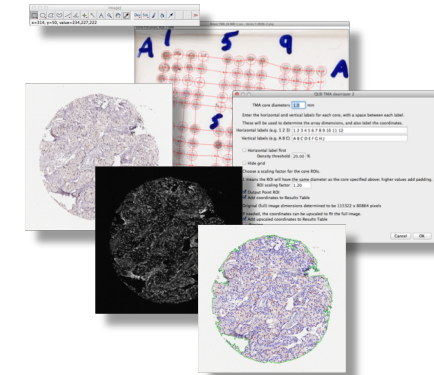
Whole slide images are very big & very complex



Most bioimage analysis tools don't support whole slide images




Attempt #1
Python scripts + MPI



Attempt #2
ImageJ plugins

Without the right tools, analysis methods are slow to create, hard to apply, not transferable... and not very useful

QuPath: Open source, whole slide image analysis software



QuPath
QuPath - Open Source Digital Pathology

[Download QuPath](#)

[Documentation](#)

[Source code](#)

Hosted on GitHub Pages — Theme by [orderedlist](#)

QuPath is open source software

QuPath aims to help improve the pathology analysis and biomarker flexible, extensible software platform.

QuPath has been developed for research & Cell Biology at Queen's University Belfast, Belfast, Northern Ireland, UK. Projects funded by Invest Northern Ireland.

Features

Whole slide viewing

Fast, flexible image viewer capable of viewing large (up to 40 GB uncompressed) using dynamic tracking slide navigation

Biomarker quantification

Received: 20 July 2017
Accepted: 21 November 2017
Published online: 04 December 2017

SCIENTIFIC REPORTS

OPEN

QuPath: Open source software for digital pathology image analysis

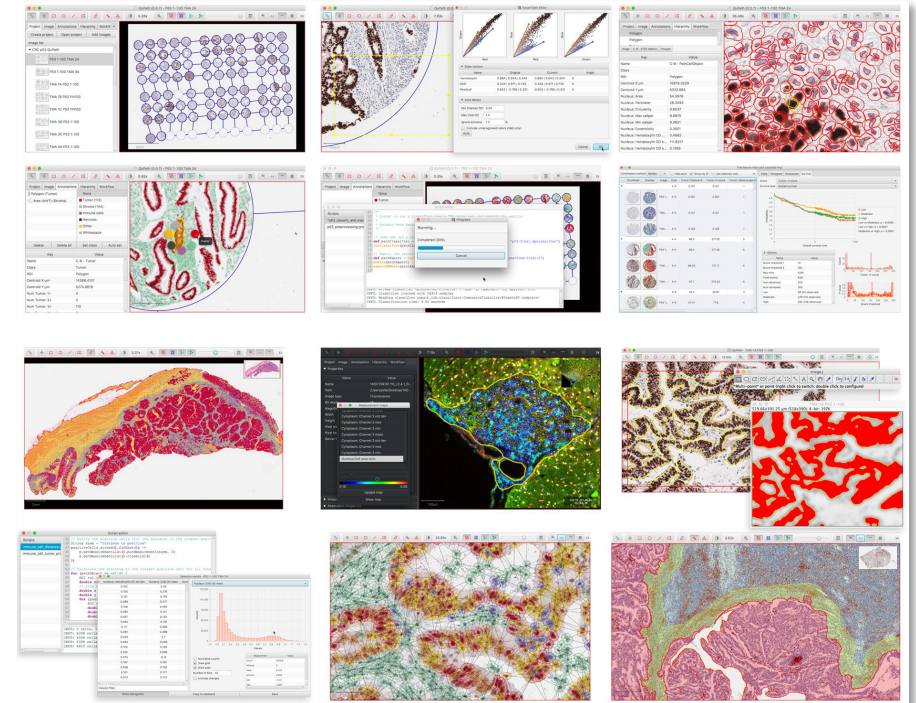
Peter Bankhead¹, Maurice B. Loughrey^{1,2}, José A. Fernández¹, Yvonne Dombrowski¹, Darragh G. McArt¹, Philip D. Dunne¹, Stephen McQuaid¹, Ronan T. Gray¹, Liam J. Murray¹, Helen G. Coleman¹, Jacqueline A. James^{1,2}, Manuel Salto-Tellez^{1,2} & Peter W. Hamilton¹

QuPath is new bioimage analysis software designed to meet the growing need for a user-friendly, extensible, open-source solution for digital pathology and whole slide image analysis. In addition to offering a comprehensive panel of tumor identification and high-throughput biomarker evaluation tools, QuPath provides researchers with powerful batch-processing and scripting functionality, and an extensible platform with which to develop and share new algorithms to analyze complex tissue images. Furthermore, QuPath's flexible design makes it suitable for a wide range of additional image analysis applications across biomedical research.

The ability to acquire high resolution digital scans of entire microscopic slides with high-resolution whole slide scanners is transforming tissue biomarker and companion diagnostic discovery through digital image analytics, automation, quantitation and objective screening of tissue samples. This area has become widely known as digital pathology^{1,2}. Whole slide scanners can rapidly generate ultra-large 2D images or z-stacks in which each plane may contain up to 40 GB uncompressed data. Manual subjective scoring of this data by traditional pathologist assessment is no longer sufficient to support large-scale tissue biomarker trials, and cannot ensure the high quality, reproducible, objective analysis essential for reliable clinical correlation and candidate biomarker selection. New and powerful software tools are urgently required to ensure that pathological assessment of tissue is practical, accessible and reliable for biological discovery and the development of clinically relevant tissue diagnostics.

In recent years, a vibrant ecosystem of open source bioimage analysis software has developed. Led by ImageJ³, researchers in multiple disciplines can now choose from a selection of powerful tools, such as Fiji⁴, Icy⁵, and CellProfiler⁶, to perform their image analyses. These open source packages encourage users to engage in further development and sharing of customized analysis solutions in the form of plugins, scripts, pipelines or workflows – enhancing the quality and reproducibility of research, particularly in the fields of microscopy and high content imaging. This template for open-source development of software has provided opportunities for image analysis to add considerably to translational research by enabling the development of the bespoke analytical methods required to address specific and emerging needs, which are often beyond the scope of existing commercial applications⁷. However, none of the aforementioned software applications tackle the specific visualization and computational challenges posed by whole slide images (WSI) and very large 2D data. Rather, open source tools for digital pathology to date have comprised libraries to handle digital slide formats (e.g. OpenSlide⁸, Bio-Formats⁹), software to crop whole slide images into manageable tiles or perform analysis on such cropped tiles (e.g. SlideToolkit¹⁰, Immunolite¹¹), or web platforms for data management and collaborative analysis (e.g. Cytomine¹²). While each of this makes a valuable contribution, the field continues to lack a commonly-accepted, open software framework for developing and distributing novel digital pathology algorithms in a manner that is immediately accessible for any researcher or pathologist. In practice, this has meant that users without access to expensive commercial solutions have had to either resort to inefficient workarounds (such as image downsampling and cropping) to apply limited quantitative analysis using general open source analysis tools to a subset of their data^{13,14}, or to rely primarily on laborious manual evaluation of slides, which is known to have high variability and limited reproducibility^{15,16}. It has also made it more difficult for computational researchers to innovate in algorithm development, and to make state-of-the-art analysis methods widely available¹⁷.

¹Northern Ireland Molecular Pathology Laboratory, Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, Northern Ireland, UK. ²Tissue Pathology, Belfast Health and Social Care Trust, Belfast, Northern Ireland, Northern Ireland, UK. ³Centre for Experimental Medicine, Queen's University Belfast, Belfast, Northern Ireland, UK. ⁴Cancer Epidemiology and Health Services Research Group, Centre for Public Health, Queen's University Belfast, Belfast, Northern Ireland, UK. Correspondence and requests for materials should be addressed to M.S.-T. (email: m.salto-tellez@qub.ac.uk) or P.W.H. (email: p.hamilton@qub.ac.uk)



Some lessons learned...

There's quite a lot involved in creating a new software platform...

Interactive whole slide image viewer Interactive drawing tools
Scanner image format support Region of interest representation
Image caching Dearranging Tissue Microarrays Object representation
Cell segmentation Feature computation
Quantification of nuclear, cytoplasmic & membranous staining Tumour identification
Membrane detection Machine learning Visualization of results
Stain estimation Texture computation Data management
Digital stain separation User interface design
Efficient data structures Command logging Help documentation
Image tiling Automation Statistical analysis
Import / Export Scripting support
Multidimensional data support Integration of survival data

A research paper could focus on one or two of these

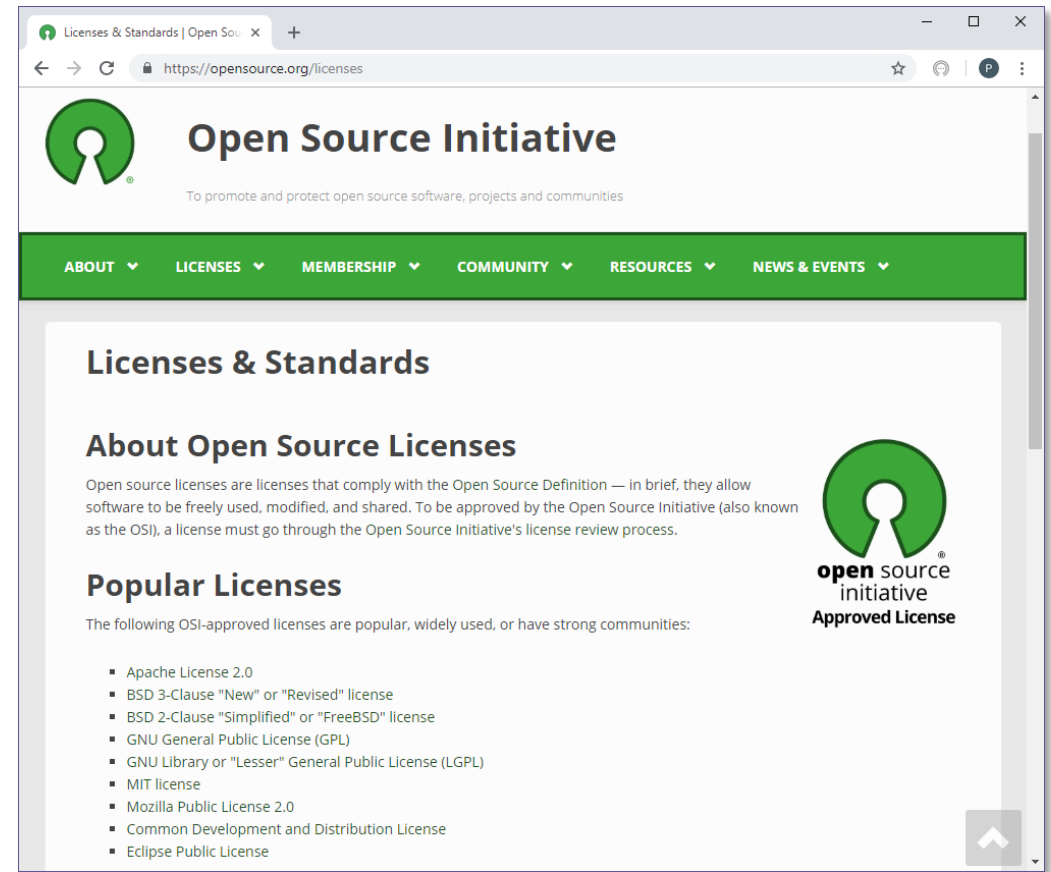
Open source can be scary



Open source can be complicated



The license choice matters



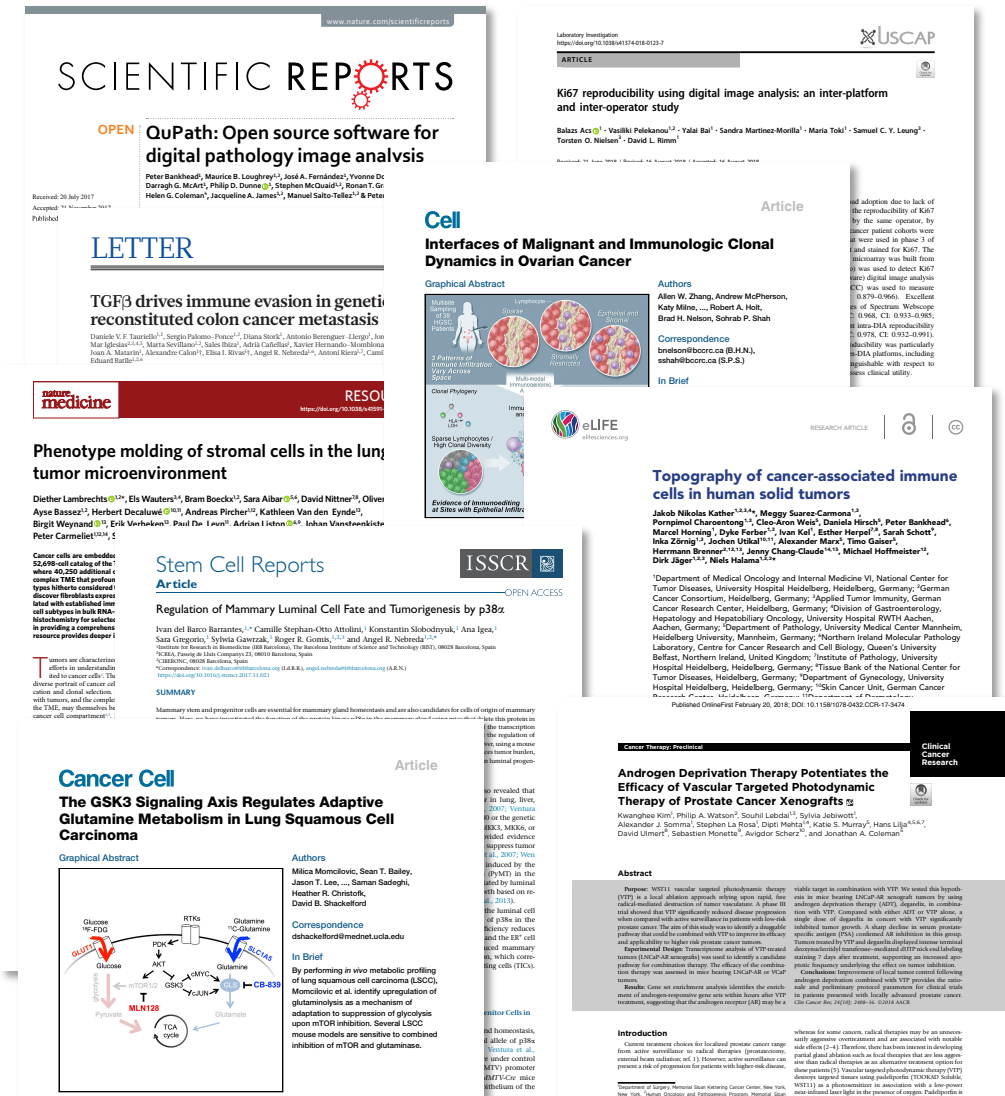
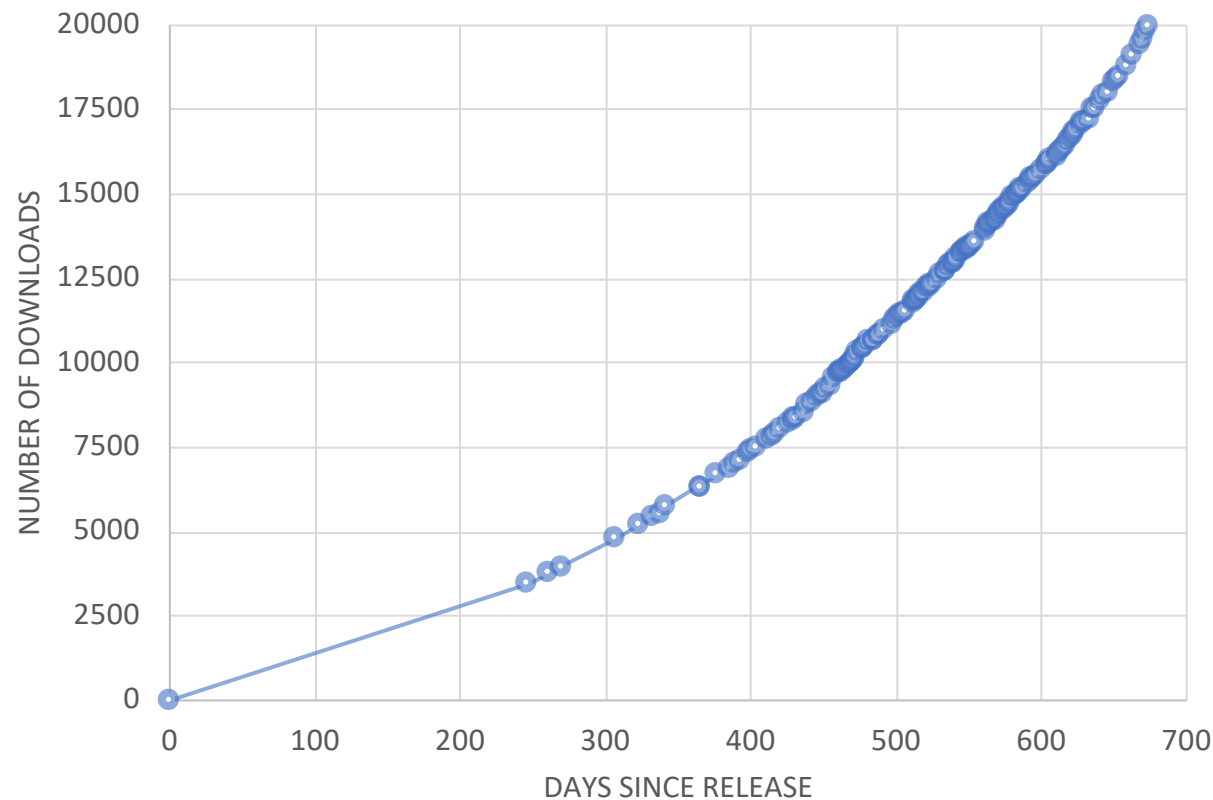
<https://opensource.org/licenses>

Open source can be costly

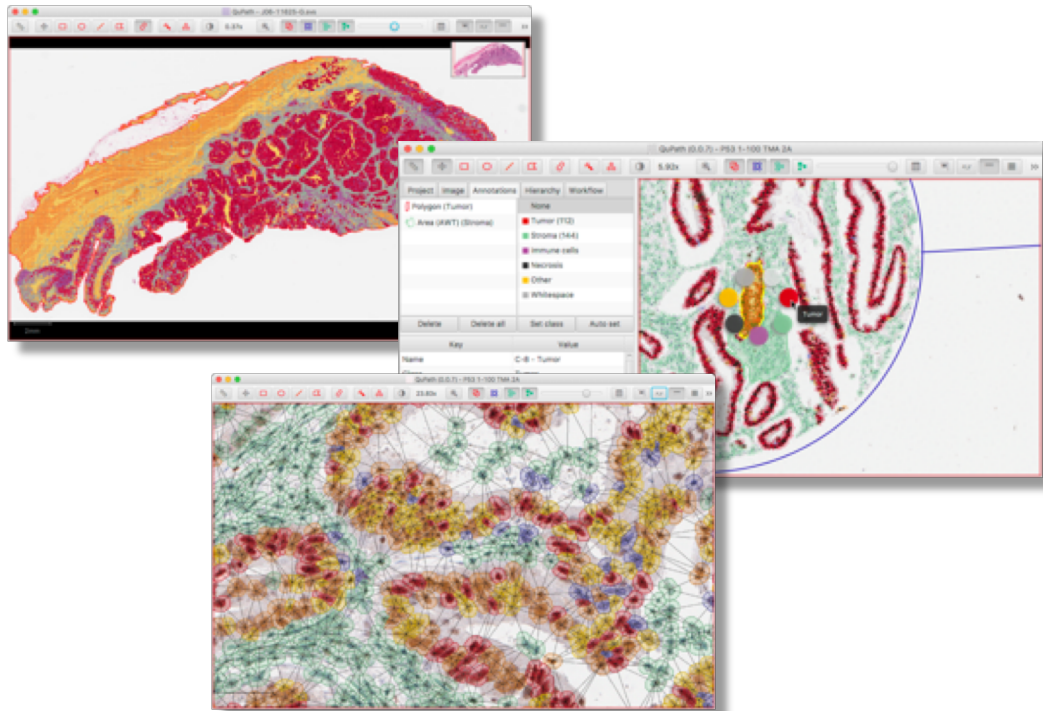


Open source can be transformative

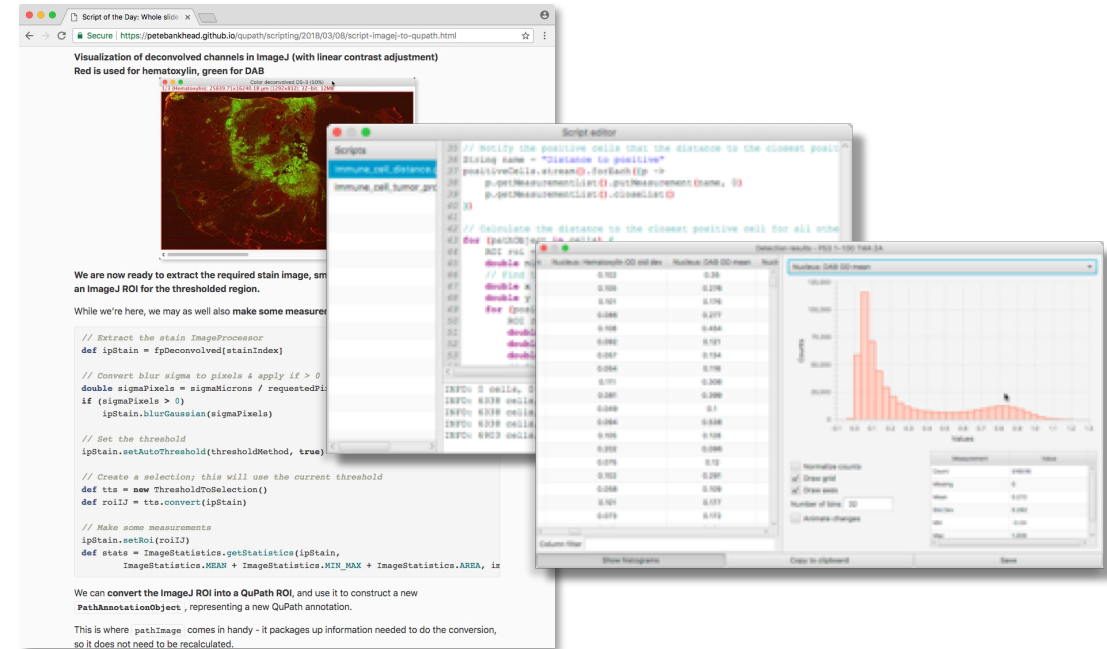
QuPath v0.1.2 downloads



User-friendliness & developer-friendliness are *both* important



Visualizations are key for
cross-disciplinary collaboration



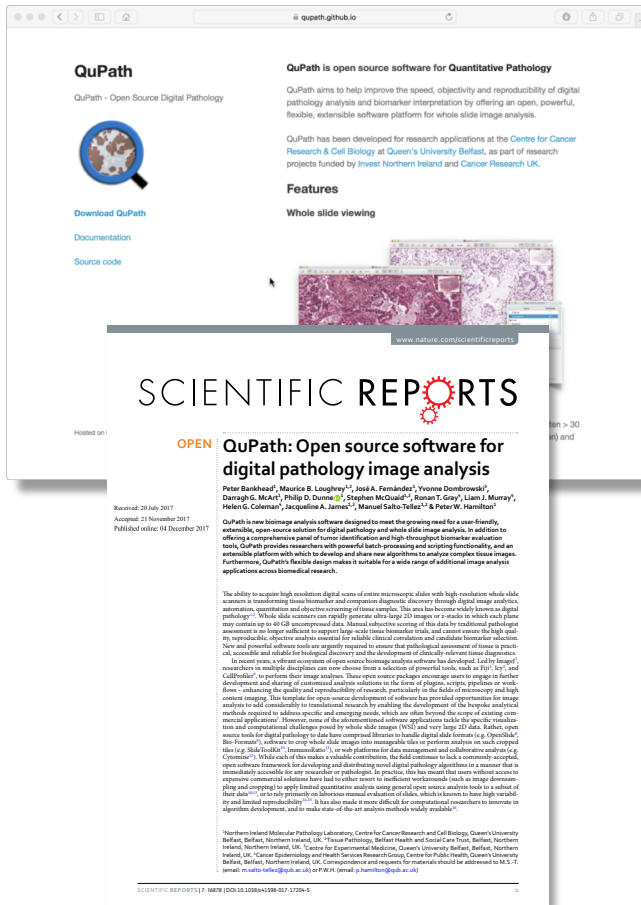
Extensibility & scripting support
make far more things possible

Reflections & Opinions

- Developing & maintaining new software can take a long time
 - But the impact it makes can *vastly* outweigh the time & cost of development
- Creating software in academia is harder than it should be
 - 'Is it really *research*? Shouldn't companies be doing it? Should we start a company...?'
 - If you use research software, *please at least cite the papers!*
- Open science is difficult & scary in practice... but absolutely necessary
 - *If* bioimaging data & analyses were fully available & transparent, I suspect we'd find
 - *Lots of poor-quality data ends up in high-quality journals*
 - *Analysis is often wrong / answers a subtly different question from what the authors thought*
 - *There are many new insights to be had using existing data*



Thanks for listening!



QuPath is software for
whole slide image analysis...



...that is freely
available &
pathologist-friendly
...



...and isn't locked down with
fixed algorithms...



...but is open source, extensible &
scriptable for developers...



...and has a wide user base
sharing ideas & insights...



<https://qupath.github.io/>